OM nucleic - nucleic search, using sw model

Copyright (c) 1993 - 2003 Compugen Ltd.

GenCore version 5.1.6

Run on: November 17, 2003, 17:30:02 ; Search time 591.3 Seconds
                                         (without alignments)
                                         1383.718 Million cell updates/sec

Title: US-09-827-383B-3
Perfect score: 20
Sequence: 1 caccgagctaatagttaag 20

Scoring table:- IDENTITY NUC
            Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database : GenEmbl:*
      1: gb_ba:*
      2: gb_htg:*
      3: gb_in:*
      4: gb_om:*
      5: gb_ov:*
      6: gb_pat:*
      7: gb_ph:*
      8: gb_pl:*
      9: gb_pri:*
     10: gb_ro:*
     11: gb_sts:*
     12: gb_sy:*
     13: gb_un:*
     14: gb_vi:*
     15: em_ba:*
     16: em_fun:*
     17: em_hum:*
     18: em_in:*
     19: em_mu:*
     20: em_om:*
     21: em_or:*
     22: em_ov:*
     23: em_pat:*
     24: em_ph:*
     25: em_pl:*
     26: em_ro:*
     27: em_sts:*
     28: em_un:*
     29: em_vi:*
     30: em_htg_hum:*
     31: em_htg_inv:*
     32: em_htg_other:*
     33: em_htg_mus:*
     34: em_htg_pln:*
     35: em_htg_rod:*
     36: em_htg_mam:*
     37: em_htg_vrt:*
     38: em_sy:*
     39: em_htgo_hum:*
     40: em_htgo_mus:*
     41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID		Description
c	1	16.8	84.0	10822	1	AE004517	AE004517	Pseudomon
	2	16.8	84.0	76730	8	AP004533	AP004533	Lotus jap
c	3	16.8	84.0	161261	9	CNS0STE3	AL358332	Human chr
	4	16.8	84.0	220760	2	AC097210	AC097210	Rattus no
c	5	16.8	84.0	231901	2	AC123417	AC123417	Rattus no
	6	16.8	84.0	240009	2	AC096200	AC096200	Rattus no
c	7	16.4	82.0	113164	2	AC136925	AC136925	Oryza sat
	8	16.4	82.0	177887	2	AC134256	AC134256	Oryza sat
c	9	16.4	82.0	217841	2	AC095995	AC095995	Rattus no
c	10	16.4	82.0	241206	2	AC120484	AC120484	Rattus no
	11	15.8	79.0	802	11	BV032188	BV032188	S212P6348
	12	15.8	79.0	41125	10	AC002108	AC002108	Genomic s
c	13	15.8	79.0	71516	9	AF117830	AF117830	Homo sapi
	14	15.8	79.0	89350	8	ATT12H17	AL021635	Arabidops
c	15	15.8	79.0	108612	10	AL844490	AL844490	Mouse DNA
c	16	15.8	79.0	146640	2	AC114335	AC114335	Canis fam
	17	15.8	79.0	161053	9	AL356608	AL356608	Human DNA
c	18	15.8	79.0	169865	9	AC079967	AC079967	Homo sapi
c	19	15.8	79.0	172071	9	AC012676	AC012676	Homo sapi
c	20	15.8	79.0	177726	9	AC004083	AC004083	Homo sapi
	21	15.8	79.0	181835	9	AC091152	AC091152	Homo sapi
	22	15.8	79.0	186608	10	AL844178	AL844178	Mouse DNA
	23	15.8	79.0	189043	8	AP004366	AP004366	Oryza sat
	24	15.8	79.0	195007	9	AC137037	AC137037	Rattus no
	25	15.8	79.0	197775	9	AC012366	AC012366	Homo sapi
	26	15.8	79.0	199577	8	ATCHRIV57	AL161557	Arabidops
c	27	15.8	79.0	201386	2	AC135720	AC135720	Mus muscu
	28	15.8	79.0	207304	2	AC107146	AC107146	Rattus no
	29	15.8	79.0	210688	10	AC127371	-AC127371	Mus muscu
c	30	15.8	79.0	214556	2	AC128105	AC128105	Rattus no
c	31	15.8	79.0	218077	2	AC097219	AC097219	Rattus no
	32	15.8	79.0	234379	2	AC109379	AC109379	Rattus no
	33	15.8	79.0	238106	2	AC123207	AC123207	Rattus no
c	34	15.8	79.0	238106	2	AC123207	AC123207	Rattus no
	35	15.8	79.0	239945	2	AC094050	AC094050	Rattus no
	36	15.8	79.0	248254	2	AC133448	AC133448	Rattus no
c	37	15.8	79.0	255270	2	AC130767	AC130767	Rattus no
	38	15.4	77.0	12375	2	AF425233	AF425233	Yersinia
	39	15.4	77.0	74925	8	AP006434	AP006434	Lotus jap
	40	15.4	77.0	132092	2	AC135422	AC135422	Oryza sat
c	41	15.4	77.0	233851	2	AC097870	AC097870	Rattus no
	42	15.4	77.0	270707	6	BX323558	BX323558	Danio rer
	43	15.2	76.0	298	6	AR250541	AR250541	Sequence
c	44	15.2	76.0	507	8	AF222834	AF222834	Phaeocisar
c	45	15.2	76.0	659	9	AJ338788	HSA338788	Homo sapi

ALIGNMENTS

RESULT 1
AE004517/c
LOCUS   AE004517   10822 bp   DNA   linear   BCT 19-FEB-2003
DEFINITION Pseudomonas aeruginosa PAO1, section 78 of 529 of the complete
            genome.
ACCESSION   AE004517 AE004091
VERSION   AE004517.1 GI:9946710
KEYWORDS   .
SOURCE   Pseudomonas aeruginosa PAO1
ORGANISM   Pseudomonas aeruginosa PAO1
            Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
            Pseudomonadaceae; Pseudomonas.
REFERENCE   1 (bases 1 to 10822)
AUTHORS   Stover,C.K., Pham,X.-Q.T., Erwin,A.L., Mizoguchi,S.D., Warrener,P.,
            Hickey,M.J., Brinkman,F.S.L., Hufnagle,W.O., Kowalik,D.J.,

```
TITLE       Lagrou,M., Garber,R.L., Goltry,L., Tolentino,E.,
            Westbrock-Wadman,S., Yuan,Y., Brody,L.L., Coulter,S.N.,
            Folger,K.R., Kas,A., Larbig,K., Lim,R.M., Smith,K.A., Spencer,D.H.,
            Wong,G.K.-S., Wu,Z. and Paulsen,I.T.
            Complete genome sequence of Pseudomonas aeruginosa PA01, an
            opportunistic pathogen
JOURNAL     Nature 406 (6799), 959-964 (2000)
MEDLINE     20437337
PUBMED      1098043
REFERENCE   2  (bases 1 to 10822)
AUTHORS     Stover,C.K., Pham,X.-Q.T., Erwin,A.L., Mizoguchi,S.D., Warrener,P.,
            Hickey,M.J., Brinkman,F.S.L., Hufnagle,W.O., Kowalik,D.J.,
            Lagrou,M., Garber,R.L., Goltry,L., Tolentino,E.,
            Westbrock-Wadman,S., Yuan,Y., Brody,L.L., Coulter,S.N.,
            Folger,K.R., Kas,A., Larbig,K., Lim,R.M., Smith,K.A., Spencer,D.H.,
            Wong,G.K.-S., Wu,Z., Paulsen,I.T., Reizer,J., Saier,M.H.,
            Hancock,R.E.W, Lory,S. and Olson,M.V.
TITLE       Direct Submission
JOURNAL     Submitted (16-MAY-2000) Department of Medicine and Genetics,
            University of Washington Genome Center, University of Washington,
            Box 352145, Seattle, WA 98195, USA
REFERENCE   3  (bases 1 to 10822)
AUTHORS     
CONSRTM     Pseudomonas aeruginosa Community Annotation Project (PseudoCAP)
TITLE       Direct Submission
JOURNAL     Submitted (04-FEB-2003) Department of Molecular Biology and
            Biochemistry, Simon Fraser University, 8888 University Dr.,
            Burnaby, British Columbia V5A 1S6, Canada
COMMENT     This represents the February 3, 2003 version of the continually
            updated, reviewed, Pseudomonas aeruginosa PA01 genome annotation,
            from PseudoCAP (see http://www.pseudomonas.com for latest updates
            and links to alternate annotation). PseudoCAP is coordinated by
            Fiona S.L. Brinkman (Simon Fraser University, Canada) and Robert
            E.W. Hancock (University of British Columbia, Canada). We welcome
            submission through www.pseudomonas.com of any proposed changes.

            'Protein name confidence' is used to rate our confidence of the
            accuracy of the protein name.
            Class 1: Function experimentally demonstrated in P. aeruginosa.
            Class 2: Function of highly similar gene experimentally
            demonstrated in another organism (and gene context consistent
            in terms of pathways its involved in, if known).
            Class 3: Function proposed based on presence of conserved amino
            acid motif, structural feature or limited sequence similarity
            to an experimentally studied gene.
            Class 4: Homologs of previously reported genes of unknown function,
            or no similarity to any previously reported sequences.
FEATURES             Location/Qualifiers
     source          1..10822
                     /organism="Pseudomonas aeruginosa PA01"
                     /mol_type="genomic DNA"
                     /strain="PA01"
                     /db_xref="taxon:208964"
     gene            81..506
                     /locus_tag="PA0817"
     CDS             81..506
                     /locus_tag="PA0817"
                     /note="Protein name confidence: Class 3 (function proposed
                     based on presence of conserved amino acid motif,
                     structural feature or limited sequence similarity to an
                     experimentally studied gene)"
                     /codon_start=1
                     /transl_table=11
                     /product="probable ring-cleaving dioxygenase"
                     /protein_id="AAG04206.1"
                     /db_xref="GI:9946711"
                     /translation="MSLMPPHLAIPVHDLPAARRFYGEVFGLSEGRSAEHWVDFDFFG
                     HQLVIHQHPQTDSQRHAGSNPVDGHGHDVPVPHFGVVLAWDDWHALAERLQQRGTRFVIE
                     PYIRFKGQVGEQATLFLFDPCGNALEFKSFRDMGQLFAK"
     gene            complement(525..797)
                     /locus_tag="PA0818"
```

```
     CDS             complement(525..797)
                     /locus_tag="PA0818"
                     /note="Protein name confidence: Class 4 (homologs of
                     previously reported genes of unknown function, or no
                     similarity to any previously reported sequences)"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAG04207.1"
                     /db_xref="GI:9946712"
                     /translation="MPYRVLLILALTPALASSATLDERLGDLEHAFKPDLDGTTSAK
                     PASPPPERFKLAPLGSVVRKKECATSSLPDBQSSCAALEFCTDCAQ"
     gene            953..1246
                     /locus_tag="PA0819"
     CDS             953..1246
                     /note="Protein name confidence: Class 4 (homologs of
                     previously reported genes of unknown function, or no
                     similarity to any previously reported sequences)"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAG04208.1"
                     /db_xref="GI:9946713"
                     /translation="MAPCPGKTVTAQSPSTFIPLRRTSQCNQHPGCATLYLDCSAASE
                     DLFDTARTRLESAIGTLSVLDLLDETEEPGESTCRITGSLLSLFEAEYARRHR"
     gene            1545..2357
                     /locus_tag="PA0820"
     CDS             1545..2357
                     /note="Protein name confidence: Class 4 (homologs of
                     previously reported genes of unknown function, or no
                     similarity to any previously reported sequences)"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAG04209.1"
                     /db_xref="GI:9946714"
                     /translation="MPAVIAENDVSIWDDETGAVYHYPRRYSAILKPGVGVVYYKGKL
                     TDKSFADQRLSAQPHYFGIARIGKVYPDPESSKGDLFALIESFQPFEQAVLAKEGNSY
                     LEPIPPSKASNYWRSGVREIGQVTFDSILARABIKPLQVAESWGDRALGGFESRTBGS
                     LSTYGVRYERDPTLRLQAIAIHGVKACGFDFGKAYGYAKGFIHVHTQPISBFE
                     APKAVNPETDLVPLCANCHAVVHRRRDRTLSVDELKGLVRGRWVCAEADVVV"
     gene            complement(2464..3552)
                     /locus_tag="PA0821"
     CDS             complement(2464..3552)
                     /locus_tag="PA0821"
                     /note="Protein name confidence: Class 4 (homologs of
                     previously reported genes of unknown function, or no
                     similarity to any previously reported sequences)"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAG04210.1"
                     /db_xref="GI:9946715"
                     /translation="MEQPFDLAAELAKQPHLLEIAGNLLMKSGPEDYIGAVLCLRGTL
                     YFKKAHTPLVRESLCQCFDBFERLAEPHLTWLWREEPAQGKPLTAYRDTQPLREMMGA
                     MDEDDHLSFCYTSGKKSRDAGAWLFDIYGKRSWQAKMGHDLSVLEFSVPLLYQERQPL
                     DFLQLFIDFARRLEPEQGYAGHAYNLSPTSWDNDBPSEAFMAARMPGLDGVTACLLAN
                     TPEFKPTRIKTVSWLTLLNNERLALAGGLDALRAQLPSSHFAFYRYGDGVVIQAGAYP
                     YIAGDAEDSRPAPYVLLNHALGIRYETIGSLHCGSSHDGELRLVGWAADQWLKRLDVE
                     DSEIPRWCDKLLSAEPYLDATNTLPERL"
     gene            complement(3569..4015)
                     /locus_tag="PA0822"
     CDS             complement(3569..4015)
                     /locus_tag="PA0822"
                     /note="Protein name confidence: Class 4 (homologs of
                     previously reported genes of unknown function, or no
                     similarity to any previously reported sequences)"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAG04211.1"
```

```
TITLE      Direct Submission
JOURNAL    Submitted (13-DEC-2001) Yasukazu Nakamura, Kazusa DNA Research
           Institute, Department of Plant Gene Research; 1532-3, Yana,
           Kisarazu, Chiba 292-0812, Japan (E-mail:ynakamu@kazusa.or.jp,
           URL:http://www.kazusa.or.jp, Tel:81-438-52-3934,
           Fax:81-438-52-3935,
FEATURES           Location/Qualifiers
     source        1..76730
                   /organism="Lotus japonicus"
                   /mol_type="genomic DNA"
                   /db_xref="taxon:34305"
                   /chromosome="3"
                   /clone_lib="LjT library"
                   /note="TAC clone:TM0080"
BASE COUNT    25039 a  13819 c  14263 g  23609 t
ORIGIN

Query Match          84.0%;  Score 16.8;  DB 8;  Length 76730;
Best Local Similarity  90.0%;  Pred. No. 1.7e+02;
Matches  18;  Conservative  0;  Mismatches  2;  Indels  0;  Gaps  0;

Qy      1 CACCGACGCTAATAGTTAAG 20
          |||||||||||||| |||||
Db  16636 CACCGACGCTAATATTTTAG 16655

RESULT 3
CNS05TE3/c
LOCUS       CNS05TE3              161261 bp    DNA     linear   PRI 13-JUL-2001
DEFINITION  Human chromosome 14 DNA sequence BAC R-255G12 of library RPCI-11
            from chromosome 14 of Homo sapiens (Human), complete sequence.
ACCESSION   AL358332
VERSION     AL358332.3  GI:14787750
KEYWORDS    HTG; HTGS_ACTIVEFIN.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 161261)
  AUTHORS   Heilig,R., Petit,J.L., Vico,V., Dasilva,C., Robert,C., Wincker,P.,
            Brottier,P., Cattolico,L., Barbe,V., Pelletier,E., Artiguenave,F.,
            Levy,M., Eckenberg,R., Bruls,T., deBerardinis,V., Cruaud,C.,
            Gyapay,G., Saurin,W. and Weissenbach,J.
  TITLE     Sequencing of the human chromosome 14
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 161261)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-JUL-2001) Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     On Jul 17, 2001 this sequence version replaced gi:10432564.
            -------------- Genome Center
            Center: Genoscope / Centre National de Sequencage
            Center code: GS
            Web site: http://www.genoscope.cns.fr/
            Contact: SeqRef@genoscope.cns.fr
            --------------
            The following BAC sequence is oriented from the T7 to the SP6 end.
            Upstream BAC (overlapping the T7 end) : R-104701S
            Downstream BAC (overlapping the SP6 end) : R-102G14 (AC=AL121125)
            -------------- Summary Statistics
            Assembly program: Phrap; version 2.0
            Quality coverage: 8.58x in Q20 bases; sum-of-contigs
            --------------
            Overall quality chart :
            Range  :  bases
              0    :
            1 - 9  :   2
           10 - 19 :   6
           20 - 29 :  57
```

```
                   /db_xref="GI:9946716"
                   /translation="MTKNPPQPILDSGTGNSPHGWIPGWISKYWDEDPEHPFKPGKG
                   MIRRPDVIIVQPNRRPTQDNIKQVVEMKFPPDPHNREQLEBDYAAIAGNKNKIVEMKP
                   SDCDCQENQRSKVPVEQAGWAVAIAGGVMFVLTRGRSPRPMIPAY"
     gene          complement(4037..4270)
                   /locus_tag="PA0823"
     CDS           complement(4037..4270)
                   /locus_tag="PA0823"
                   /note="Protein name confidence: Class 4 (homologs of
                   previously reported genes of unknown function, or no
                   similarity to any previously reported sequences)"
                   /codon_start=1
                   /transl_table=11
                   /product="hypothetical protein"
                   /protein_id="AAG04212.1"
                   /db_xref="GI:9946717"
                   /translation="MSDQPSLSRGGMSPEGRTNPVGTLEPRLDPQDKKVLCSAVCHCS
                   STPNLSQDGKSLKQACVAQRLGELDEVLQGRSP"
     gene          complement(4267..4785)
                   /locus_tag="PA0824"
     CDS           complement(4267..4785)
                   /locus_tag="PA0824"
                   /note="Protein name confidence: Class 4 (homologs of
                   previously reported genes of unknown function, or no
                   similarity to any previously reported sequences)"
                   /codon_start=1
                   /transl_table=11
                   /product="hypothetical protein"
                   /protein_id="AAG04213.1"
                   /db_xref="GI:9946718"
                   /translation="MSGKPAARVTDPTTCPVPGHGSNPIVQGSPDVVFDGLPAARGGD
                   ASACGSPMISAVSSTVLINGLPAVTLGSIGAHGNVVIGGSGTVLIGDVFTPAPRAPAL
                   PLNRNSVPCSGRFQLIDHETGKPVAGRRVRVWSSGGWNAFDTTDADGMTSWIERPTAE
                   ILYIDLVQRCDA"
     gene          complement(4959..5294)
                   /locus_tag="PA0825"
     CDS           complement(4959..5294)
                   /locus_tag="PA0825"
                   /note="Protein name confidence: Class 4 (homologs of
                   previously reported genes of unknown function, or no
                   similarity to any previously reported sequences)"

Query Match          84.0%;  Score 16.8;  DB 1;  Length 10822;
Best Local Similarity  90.0%;  Pred. No. 1.8e+02;
Matches  18;  Conservative  0;  Mismatches  2;  Indels  0;  Gaps  0;

Qy      1 CACCGACGCTAATAGTTAAG 20
          ||| ||||||| ||||||||
Db   1355 CACAGACGCTCATAGTTAAG 1336

RESULT 2
AP004533
LOCUS       AP004533             76730 bp    DNA     linear   PLN 14-DEC-2001
DEFINITION  Lotus japonicus genomic DNA, chromosome 3, clone:LjT14G02, TM0080,
            complete sequence.
ACCESSION   AP004533
VERSION     AP004533.1  GI:17736900
KEYWORDS    HTG.
SOURCE      Lotus japonicus
  ORGANISM  Lotus japonicus
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Loteae;
            Lotus.
REFERENCE   1
  AUTHORS   Sato,S., Kaneko,T., Nakamura,Y., Asamizu,E., Kato,T. and Tabata,S.
  TITLE     Structural Analysis of a Lotus japonicus Genome. I. Sequence
            Features and Mapping of Fifty-six TAC clones which cover the 5.4 Mb
            Regions of the Genome
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 76730)
  AUTHORS   Nakamura,Y.
```

```
30 - 39    :        189
40 - 49    :       2710
50 - 59    :       8979
60 - 69    :      11203
70 - 79    :      22101
80 - 89    :      44981
90 - 99    :      71033
```

Percentage of bases with a quality value >= 40 : 99 %.

```
FEATURES             Location/Qualifiers
source               1..161261
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="14"
                     /clone_lib="R-255G12"
                     /clone_lib="RPCI-11"
STS                  126666..126895
                     /note="matching EMBL:Z52148
                     RHdb:RH42652
                     RHdb:RH31310
                     dbSTS:STS2223
                     Identified using the e-PCR software (G. Schuler)"
STS                  154125..154250
                     /note="matching EMBL:T86731
                     RHdb:RH53680
                     RHdb:RH32887
                     dbSTS:STS26145
                     Identified using the e-PCR software (G. Schuler)"
BASE COUNT   45445 a   31846 c   34019 g   49951 t
ORIGIN
```

Query Match   84.0%;   Score 16.8;   DB 9;   Length 161261;
Best Local Similarity 90.0%;   Pred. No. 1.7e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy    1 CACCGACGCTAATAGTTAAG 20
         ||||| || ||||||||||||
Db 10507 CACCAACTCTAATAGTTAAG 10488
```

RESULT 4
AC097210/c
LOCUS   AC097210   220760 bp DNA   linear   HTG 10-MAY-2003
DEFINITION Rattus norvegicus clone CH230-207E1, WORKING DRAFT SEQUENCE, 2
            unordered pieces.
ACCESSION   AC097210
VERSION   AC097210.7 GI:30521257
KEYWORDS   HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE   Rattus norvegicus (Norway rat)
ORGANISM   Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1 (bases 1 to 220760)
AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davis,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,R., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,A., Guerra,M., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,C.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,J., Johnson,B., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,K., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,J., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,J., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindarne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarmpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,D., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,R., Wei,K., White,F.,
            Williams,G., Willson,R., Wlecyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
TITLE   Direct Submission
JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 220760)
AUTHORS   Worley,K.C.
TITLE   Direct Submission
JOURNAL   Submitted (12-OCT-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3 (bases 1 to 220760)
AUTHORS   Rat Genome Sequencing Consortium.
TITLE   Direct Submission
JOURNAL   Submitted (10-MAY-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT   On May 10, 2003 this sequence version replaced gi:2494258.2.
            The sequence in this assembly is a combination of BAC based atlas
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contig within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            ---------- Genome Center
            Center: Baylor College of Medicine
            Center code: BCM
            Web site: http://www.hgsc.bcm.tmc.edu/
            Contact: hgsc-help@bcm.tmc.edu
            ---------- Project Information
            Center project name: GIMY
            Center clone name: CH230-207E1
            ---------- Summary Statistics
            Assembly program: Atlas 3.0;
            Consensus quality: 209257 bases at least Q40
            Consensus quality: 211357 bases at least Q30
            Consensus quality: 212520 bases at least Q20
            Estimated insert size: 216663; sum-of-contigs estimation
            Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
```

```
      ----------
      * NOTE: Estimated insert size may differ from sequence length
      *       (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
      * NOTE: This is a 'working draft' sequence. It currently
      *       consists of 2 contigs. The true order of the pieces
      *       is not known and their order in this sequence record is
      *       arbitrary. Gaps between the contigs are represented as
      *       runs of N, but the exact sizes of the gaps are unknown.
      *       This record will be updated with the finished sequence
      *       as soon as it is available and the accession number will
      *       be preserved.
      *
      *       1    219368: contig of 219368 bp in length
      *   219369   219468: gap of unknown length
      *   219469   220760: contig of 1292 bp in length.
FEATURES             Location/Qualifiers
      source          1..220760
                      /organism="Rattus norvegicus"
                      /mol_type="genomic DNA"
                      /db_xref="taxon:10116"
                      /clone="CH230-207E1"
      misc_feature    1..1836
                      /note="wgs end extension
                      clone end:Sp6"
      misc_feature    5729..6382
                      /note="clone_boundary
                      clone end:Sp6
                      site:EcoRI
                      end_sequence:BH289343"
      misc_feature    complement(216545..217346)
                      /note="clone_boundary
                      clone end:T7
                      site:EcoRI
                      end_sequence:RWBK125TJB"
BASE COUNT      61737 a   43269 c   43471 g   64832 t    7451 others
ORIGIN

Query Match     84.0%;  Score 16.8;  DB 2;  Length 220760;
Best Local Similarity   90.0%;  Pred. No. 1.7e+02;
Matches  18;  Conservative  0;  Mismatches  2;  Indels  0;  Gaps  0;

Qy        1 CACCGACGCTAATAGTTAAG 20
            |||| || |||||||||||||
Db   175855 CACCTACTCTAATAGTTAAG 175836

RESULT 5
AC123417
LOCUS       AC123417          231901 bp   DNA   linear   HTG 19-NOV-2002
DEFINITION  Rattus norvegicus clone CH230-195F22, WORKING DRAFT SEQUENCE.
ACCESSION   AC123417
VERSION     AC123417.4  GI:25088183
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 231901)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,K., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,D., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
```

TITLE

            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,M., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,K., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Louiseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindaratne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarrpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,R., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Wilson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,J., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
TITLE   Direct Submission
JOURNAL   Unpublished
REFERENCE 2 (bases 1 to 231901)
  AUTHORS   Worley,K.C.
TITLE   Direct Submission
JOURNAL   Submitted (29-MAY-2002) Human Genome Sequencing Center, Department
          of Molecular and Human Genetics, Baylor College of Medicine, One
          Baylor Plaza, Houston, TX 77030, USA
REFERENCE 3 (bases 1 to 231901)
  AUTHORS   Rat Genome Sequencing Consortium.
TITLE   Direct Submission
JOURNAL   Submitted (19-NOV-2002) Human Genome Sequencing Center, Department
          of Molecular and Human Genetics, Baylor College of Medicine, One
          Baylor Plaza, Houston, TX 77030, USA
COMMENT   On Nov 19, 2002 this sequence version replaced gi:23265439.
          The sequence in this sequence is a combination of BAC based reads
          and whole genome shotgun sequencing reads assembled using Atlas
          (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
          in the feature table below represents a 'scaffold' in the Atlas
          assembly (a 'contig-scaffold'). Within each contig-scaffold,
          individual sequence contigs are ordered and oriented and separated
          by sized gaps filled with Ns to the estimated size. The sequence
          may extend beyond the ends of the clone and there may be sequence
          contigs within a contig-scaffold that consist entirely of whole
          genome shotgun sequence reads. Both end sequences and whole genome
          shotgun sequence only contigs will be indicated in the feature
          table.
          --------------- Genome Center
          Center: Baylor College of Medicine
          Center code: BCM
          Web site: http://www.hgsc.bcm.tmc.edu/
          Contact: hgsc-help@bcm.tmc.edu
          --------------- Project Information
          Center project name: GXDR
          Center clone name: CH230-195F22
          --------------- Summary Statistics
          Assembly program: Phrap; version 0.990329
          Consensus quality: 212992 bases at least Q40
          Consensus quality: 215592 bases at least Q30

```
        Consensus quality: 216777 bases at least Q20
        Estimated insert size: 216147; sum-of-contigs estimation
        Quality coverage: 7x in Q20 region; sum-of-contigs estimation
        -----------------
      * NOTE: Estimated insert size may differ from sequence length
      *      (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
      * NOTE: This is a 'working draft' sequence. It currently
      * consists of 1 contigs. Gaps between the contigs
      * are represented as runs of N. The order of the pieces
      * is believed to be correct as given, however the sizes
      * of the gaps between them are based on estimates that have
      * provided by the submittor.
      * This sequence will be replaced
      * by the finished sequence as soon as it is available and
      * the accession number will be preserved.

FEATURES             Location/Qualifiers
     source          1..231901
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-195F22"
     misc_feature    1..1092
                     /note="wgs end extension
                     clone_end:T7"
     misc_feature    2667..4220
                     /note="wgs end extension
                     clone_end:T7"
     misc_feature    4271..5756
                     /note="wgs end extension
                     clone_end:T7"
     misc_feature    6046..6678
                     /note="clone_boundary
                     clone_end:T7
                     site:
                     end sequence:BH343171"
     misc_feature    70625..71968
                     /note="wgs contig"
     misc_feature    complement(224249..224798)
                     /note="clone_boundary
                     clone_end:Sp6
                     site:
                     end sequence:BH343212"
     misc_feature    230060..231901
                     /note="wgs end extension
                     clone_end:Sp6"

BASE COUNT   58062 a   48279 c   49404 g   61493 t   14663 others
ORIGIN

Query Match             84.0%;  Score 16.8;  DB 2;  Length 231901;
Best Local Similarity   90.0%;  Pred. No. 1.7e+02;
Matches   18;  Conservative  0;  Mismatches  2;  Indels  0;  Gaps  0;

Qy      1 CACCGACGCTAATAGTTAAG 20
          ||| ||||||||||||||||
Db 130601 CACACACGCTAATAGTTAAG 130620
```

RESULT 6
AC096200

```
LOCUS       AC096200              240009 bp   DNA    linear   HTG 10-MAY-2003
DEFINITION  Rattus norvegicus clone CH230-37K11, WORKING DRAFT SEQUENCE, 3
            unordered pieces.
ACCESSION   AC096200
VERSION     AC096200.6  GI:30522427
KEYWORDS    HTG; HTGS_PHASE1; HTGS DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 240009)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswal,O.K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,J., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,S., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,Y., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorenshewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newon,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,J., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,E., Sodergren,E., Song,J.-Z., Sorelle,A., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,V., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,X., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 240009)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-SEP-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine. One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 240009)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-MAY-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine. One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On May 10, 2003 this sequence version replaced gi:24818305.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
```

```
            shotgun sequence only contigs will be indicated in the feature
            table.
            ------------ Genome Center
            Center: Baylor College of Medicine
            Center code: BCM
            Web site: http://www.hgsc.bcm.tmc.edu/
            Contact: hgsc-help@bcm.tmc.edu
            ------------ Project Information
            Center project name: GEOL
            Center clone name: CH230-37K11
            ------------ Summary Statistics
            Assembly program: Atlas 3.0;
            Consensus quality: 226448 bases at least Q40
            Consensus quality: 228877 bases at least Q30
            Consensus quality: 230092 bases at least Q20
            Estimated insert size: 233615; sum-of-contigs estimation
            Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

        * NOTE: Estimated insert size may differ from sequence length
        *       (see http://www.hgsc.bcm.tmc.edu/docs/Genbank draft data.html).
        * NOTE: This is a 'working draft' sequence. It currently
        *       consists of 3 contigs. The true order of the pieces
        *       is not known and their order in this sequence record is
        *       arbitrary. Gaps between the contigs are represented as
        *       runs of N, but the exact sizes of the gaps are unknown.
        *       This record will be updated with the finished sequence
        *       as soon as it is available and the accession number will
        *       be preserved.
        *         1  237591: contig of 237591 bp in length
        *    237592  237691: gap of unknown length
        *    237692  238737: contig of 1046 bp in length
        *    238738  238837: gap of unknown length
        *    238838  240009: contig of 1172 bp in length

FEATURES             Location/Qualifiers
     source          1..240009
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-37K11"
     misc_feature    1..1403
                     /note="wgs_end_extension
                     clone_end:T7"
     misc_feature    5429..7521
                     /note="wgs_end_extension
                     clone_end:T7"
     misc_feature    7557..8401
                     /note="clone_boundary
                     clone_end:T7
                     site:EcoRI
                     end_sequence:BH305843"
     misc_feature    141219..225407
                     /note="clone_boundary
                     clone_end:Sp6
                     site:EcoRI
                     end_sequence:BH305845"

BASE COUNT    62638 a   49403 c   51000 g   68223 t    8745 others
ORIGIN

Query Match    84.0%;  Score 16.8;  DB 2;  Length 240009;
Best Local Similarity 90.0%;  Pred. No. 1.7e+02;
Matches    18;  Conservative  0;  Mismatches  2;  Indels  0;  Gaps  0;

Qy      1 CACCGACGCTAATAGTTAAG 20
          ||||||| |||||||||||||
Db 220049 CACACACGCTAATAGTTAAG 220068

RESULT 7
AC136925/c
LOCUS       AC136925      113164 bp    DNA     linear   HTG 11-NOV-2002
DEFINITION  Oryza sativa (japonica cultivar-group) chromosome 11 clone
            OSJNBa0090B19, *** SEQUENCING IN PROGRESS ***, 7 ordered pieces.
ACCESSION   AC136925
```

```
VERSION     AC136925.1  GI:24850539
KEYWORDS    HTG; HTGS PHASE2.
SOURCE      Oryza sativa (japonica cultivar-group)
ORGANISM    Oryza sativa (japonica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 113164)
  AUTHORS   Pal,A.K., Dixit,A., Sureshbabu,K., Singh,A., Pal,S., Yadav,M.,
            Gaikwad,K., Ghazi,I.A., Swain,S.C., Srivastava,S., Bhargava,A.,
            Dalal,V., Batra,K., Sharma,T.R., Mohapatra,T. and Singh,N.K.
  TITLE     Genomic sequence for Oryza sativa chromosome 11
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 113164)
  AUTHORS   Pal,A.K., Dixit,A., Sureshbabu,K., Singh,A., Pal,S., Yadav,M.,
            Gaikwad,K., Ghazi,I.A., Swain,S.C., Srivastava,S., Bhargava,A.,
            Dalal,V., Batra,K., Sharma,T.R., Mohapatra,T. and Singh,N.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-NOV-2002) IIRGS, NRC on Plant Biotechnology, Indian
            Agricultural Research Institute, LBS Centre, New Delhi, Delhi
            110012, India
COMMENT     * NOTE: This is a 'working draft' sequence. It currently
            *       consists of 7 contigs. Gaps between the contigs
            *       are represented as runs of N. The order of the pieces
            *       is believed to be correct as given, however the sizes
            *       of the gaps between them are based on estimates that have
            *       provided by the submitter.
            *       This sequence will be replaced
            *       by the finished sequence as soon as it is available and
            *       the accession number will be preserved.
            *         1    1160: contig of 1160 bp in length
            *      1161    1260: gap of unknown length
            *      1261   37630: contig of 36370 bp in length
            *     37631   37730: gap of unknown length
            *     37731   50155: contig of 12425 bp in length
            *     50156   50255: gap of unknown length
            *     50256   52672: contig of 2417 bp in length
            *     52673   52772: gap of unknown length
            *     52773   64423: contig of 11651 bp in length
            *     64424   64523: gap of unknown length
            *     64524   69586: contig of 5063 bp in length
            *     69587   69686: gap of unknown length
            *     69687  113164: contig of 43478 bp in length

FEATURES             Location/Qualifiers
     source          1..113164
                     /organism="Oryza sativa (japonica cultivar-group)"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:39947"
                     /chromosome="11"
                     /clone="OSJNBa0090B19"
                     /note="(japonica cultivar-group)"
BASE COUNT    34095 a   22866 c   22420 g   33175 t    608 others
ORIGIN

Query Match    82.0%;  Score 16.4;  DB 2;  Length 113164;
Best Local Similarity 94.4%;  Pred. No. 3e+02;
Matches    17;  Conservative  0;  Mismatches  1;  Indels  0;  Gaps  0;

Qy      2 ACCGACGCTAATAGTTAA 19
          ||||||||||||| ||||
Db  40250 ACCGACGCTAATACTTAA 40233

RESULT 8
AC134256
LOCUS       AC134256     177887 bp    DNA     linear   HTG 28-FEB-2003
DEFINITION  Oryza sativa (japonica cultivar-group) chromosome 11 clone
            OSJNBa0033H01, *** SEQUENCING IN PROGRESS ***, 7 ordered pieces.
ACCESSION   AC134256
VERSION     AC134256.3  GI:28604227
KEYWORDS    HTG; HTGS PHASE2.
SOURCE      Oryza sativa (japonica cultivar-group)
ORGANISM    Oryza sativa (japonica cultivar-group)
```

REFERENCE   1 (bases 1 to 177887)
AUTHORS   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
             Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
             Ehrhartoideae; Oryzeae; Oryza.
             Bhargava,A., Pal,S., Batra,K., Yadav,M., Sureshbabu,K., Singh,A.,
             Srivastava,S., Swain,S.C., Dixit,A., Ghazi,I.A., Dalal,V.,
             Pal,A.K., Gaikwad,K., Sharma,T.R., Mohapatra,T. and Singh,N.K.
TITLE   Genomic sequence for Oryza sativa chromosome 11
JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 177887)
AUTHORS   Bhargava,A., Pal,S., Batra,K., Yadav,M., Sureshbabu,K., Singh,A.,
             Srivastava,S., Swain,S.C., Dixit,A., Ghazi,I.A., Dalal,V.,
             Pal,A.K., Gaikwad,K., Sharma,T.R., Mohapatra,T. and Singh,N.K.
TITLE   Direct Submission
JOURNAL   Submitted (24-SEP-2002) IIRGS, NRC on Plant Biotechnology, Indian
             Agricultural Research Institute, LBS Centre, New Delhi, Delhi
             11012, India
REFERENCE   3 (bases 1 to 177887)
AUTHORS   Bhargava,A., Pal,S., Batra,K., Yadav,M., Sureshbabu,K., Singh,A.,
             Srivastava,S., Swain,S.C., Dixit,A., Ghazi,I.A., Dalal,V.,
             Pal,A.K., Gaikwad,K., Sharma,T.R., Mohapatra,T. and Singh,N.K.
TITLE   Direct Submission
JOURNAL   Submitted (28-FEB-2003) IIRGS, NRC on Plant Biotechnology, Indian
             Agricultural Research Institute, LBS Centre, New Delhi, Delhi
             11012, India
COMMENT   On Feb 28, 2003 this sequence version replaced gi:27502446.
             * NOTE: This is a 'working draft' sequence. It currently
             * consists of 7 contigs. Gaps between the contigs
             * are represented as runs of N. The order of the pieces
             * is believed to be correct as given, however the sizes
             * of the gaps between them are based on estimates that have
             * provided by the submittor.
             * This sequence will be replaced
             * by the finished sequence as soon as it is available and
             * the accession number will be preserved.
             *   10831: contig of 10831 bp in length
             *   10931: gap of unknown length
             *   10932 49727: contig of 38796 bp in length
             *   49727: gap of unknown length
             *   49728 49827: gap of unknown length
             *   49828 62445: contig of 12618 bp in length
             *   62446 62545: gap of unknown length
             *   62546 105748: contig of 43203 bp in length
             *   105749 105848: gap of unknown length
             *   105849 117930: contig of 12082 bp in length
             *   117931 118030: gap of unknown length
             *   118031 119573: contig of 1543 bp in length
             *   119574 119673: gap of unknown length
             *   119674 177887: contig of 58214 bp in length.
FEATURES   Location/Qualifiers
     source   1..177887
                     /organism="Oryza sativa (japonica cultivar-group)"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:39947"
                     /chromosome="11"
                     /clone="OSJNBa0033H01"
                     /note="japonica cultivar-group)"
BASE COUNT   52016 a 35110 c 36445 g 53713 t 603 others
ORIGIN

Query Match   82.0%; Score 16.4; DB 2; Length 177887;
Best Local Similarity 94.4%; Pred. No. 2.9e+02;
Matches   17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy   2 ACCGACGCTAATAGTTAA 19
          ||||||||||||| ||||
Db 75066 ACCGACGCTAATACTTAA 75083

RESULT 9
AC095995/c
LOCUS   AC095995   217841 bp   DNA   linear   HTG 10-MAY-2003
DEFINITION Rattus norvegicus clone CH230-32A21, WORKING DRAFT SEQUENCE, 2
            unordered pieces.
ACCESSION   AC095995
VERSION   AC095995.7 GI:30522852
KEYWORDS   HTG; HTGS_PHASE1; HTGS DRAFT; HTGS_FULLTOP.
SOURCE   Rattus norvegicus (Norway rat)
ORGANISM   Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1 (bases 1 to 217841)
AUTHORS   Muzny,D.Marie, Metzker,M.Lee, Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,R., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hernandez,J.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,R., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarmpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,J., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Wilson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausen,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
TITLE   Direct Submission
JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 217841)
AUTHORS   Worley,K.C.
TITLE   Direct Submission
JOURNAL   Submitted (17-SEP-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3 (bases 1 to 217841)
AUTHORS   Rat Genome Sequencing Consortium.
TITLE   Direct Submission
JOURNAL   Submitted (10-MAY-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT   On May 10, 2003 this sequence version replaced gi:24818121.

```
LOCUS       AC120484            241206 bp    DNA     linear   HTG 12-OCT-2002
DEFINITION  Rattus norvegicus clone CH230-135H12, WORKING DRAFT SEQUENCE.
ACCESSION   AC120484
VERSION     AC120484.3  GI:23907778
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 241206)
  AUTHORS   Muzny,D.Marie, Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Garza,M.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,D., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,J., Liu,Y., London,P., Longacre,S., Lopez,J., Liu,J.,
            Lorensuhewa,L., Louiseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarmpunsagoon,A., Pai,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,G., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,R., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,E., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,J., Warren,R., Wei,X.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wieczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhang,J., Zhou,X., Zhou,X., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 241206)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-MAY-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 241206)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-OCT-2002) Human Genome Sequencing Center, Department
```

The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.

```
       ------------ Genome Center
       Center: Baylor College of Medicine
       Center code: BCM
       Web site: http://www.hgsc.bcm.tmc.edu/
       Contact: hgsc-help@bcm.tmc.edu
       ------------ Project Information
       Center project name: GECQ
       Center clone name: CH230-32A21
       ------------ Summary Statistics
       Assembly program: Atlas 3.0;
       Consensus quality: 209555 bases at least Q40
       Consensus quality: 211996 bases at least Q30
       Consensus quality: 211996 bases at least Q30
       Consensus quality: 213711 bases at least Q20
       Estimated insert size: 219326; sum-of-contigs estimation
       Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
       ------------
*      NOTE: Estimated insert size may differ from sequence length
*            (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
*      NOTE: This is a 'working draft' sequence. It currently
*            consists of 2 contigs. The true order of the pieces
*            is not known and their order in this sequence record is
*            arbitrary. Gaps between the contigs are represented as
*            runs of N, but the exact sizes of the gaps are unknown.
*            This record will be updated with the finished sequence
*            as soon as it is available and the accession number will
*            be preserved.
*            1    216619: contig of 216619 bp in length
*       216620    216719: gap of unknown length
*       216720    217841: contig of 1122 bp in length.
FEATURES             Location/Qualifiers
     source          1..217841
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-32A21"
     misc_feature    1..1835
                     /note="wgs end extension
                     clone_end:T7"
     misc_feature    3321..130068
                     /note="clone_boundary
                     clone_end:T7
                     site:EcoRI
                     end_sequence:BH325857"
     misc_feature    complement(215092..215837)
                     /note="clone_boundary
                     clone_end:Sp6
                     site:EcoRI
                     end_sequence:BH325858"
BASE COUNT    62710 a   46664 c   45725 g   59600 t   3142 others
ORIGIN
```

```
Query Match      82.0%;  Score 16.4;  DB 2;  Length 217841;
Best Local Similarity   94.4%;  Pred. No. 2.9e+02;
Matches   17;  Conservative  0;  Mismatches  1;  Indels  0;  Gaps  0;

Qy       3 CCGACGCTAATAGTTAAG 20
             ||||| ||||||||||||
Db   37547 CCGACTCTAATAGTTAAG 37530
```

RESULT 10

COMMENT   of Molecular and Human Genetics, Baylor College of Medicine, One
             Baylor Plaza, Houston, TX 77030, USA
             On Oct 12, 2002 this sequence version replaced gi:21908466.
             The sequence in this assembly is a combination of BAC based reads
             and whole genome shotgun sequencing reads assembled using Atlas
             (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
             in the feature table below represents a scaffold in the Atlas
             assembly (a 'contig-scaffold'). Within each contig-scaffold,
             individual sequence contigs are ordered and oriented, and separated
             by sized gaps filled with Ns to the estimated size. The sequence
             may extend beyond the ends of the clone and there may be sequence
             contigs within a contig-scaffold that consist entirely of whole
             genome shotgun sequence reads. Both end sequences and whole genome
             shotgun sequence only contigs will be indicated in the feature
             table.
             ------- Genome Center
             Center: Baylor College of Medicine
             Center code: BCM
             Web site: http://www.hgsc.bcm.tmc.edu/
             Contact: hgsc-help@bcm.tmc.edu
             ------- Project Information
             Center project name: GXSF
             Center clone name: CH230-135H12
             ------- Summary Statistics
             Assembly program: Phrap; version 0.990329
             Consensus quality: 228650 bases at least Q40
             Consensus quality: 230295 bases at least Q30
             Consensus quality: 231252 bases at least Q20
             Estimated insert size: 235536; sum-of-contigs estimation
             Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
             * NOTE: Estimated insert size may differ from sequence length
             *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
             * NOTE: This is a 'working draft' sequence. It currently
             *   consists of 1 contigs. Gaps between the contigs
             *   are represented as runs of N. The order of the pieces
             *   is believed to be correct as given, however the sizes
             *   of the gaps between them are based on estimates that have
             *   provided by the submittor.
             *   This sequence will be replaced
             *   by the finished sequence as soon as it is available and
             *   the accession number will be preserved.
             *   1   241206: contig of 241206 bp in length.
FEATURES   Location/Qualifiers
     source   1..241206
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-135H12"
     misc_feature   820..1685
                     /note="clone_boundary
                     clone_end:T7
                     site:EcoRI
                     end sequence:RWAWL42TJB"
BASE COUNT   62790 a   53676 c   52951 g   62493 t   9296 others
ORIGIN

Query Match   82.0%; Score 16.4; DB 2; Length 241206;
  Best Local Similarity   94.4%; Pred. No. 2.9e+02;
  Matches   17; Conservative   0; Mismatches   1; Indels   0; Gaps   0;

Qy   3 CCGACGCCTAATAGTTAAG 20
              ||| ||||||||||||||||
Db   127576 CCGACTCTAATAGTTAAG 127559

RESULT 11
BV032188
LOCUS   BV032188   802 bp   DNA   linear   STS 30-MAY-2003
DEFINITION S212P6348FA9.T0 CZECHII/Ei Mus musculus STS genomic. sequence
            tagged site.
ACCESSION   BV032188
VERSION   BV032188.1 GI:31116083

KEYWORDS   STS.
SOURCE   Mus musculus (house mouse)
  ORGANISM Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 802)
  AUTHORS   Wade,C.M., Kulbokas,E.J. III, Kirby,A.W., Zody,M.C., Mullikin,J.C.,
            Lander,E.S., Lindblad-Toh,K. and Daly,M.J.
  TITLE   The mosaic structure of variation in the laboratory mouse genome
  JOURNAL   Nature 420 (6915), 574-578 (2002)
  MEDLINE   22354684
   PUBMED   12466852
COMMENT   Contact: Kerstin Lindblad-Toh
            Whitehead Institute for Biomedical Research, Center for Genome
            Research
            320 Charles Street, Cambridge, MA 02141, USA
            Tel: 6172521477
            Fax: 6172580903
            Email: kersli@genome.wi.mit.edu
            Primer A: None
            Primer B: None
            STS size: 802
            Protocol:
            WGS-discovery: Paired-end low-coverage whole genome shotgun reads
            were generated from 129S1/SvImJ, C3H/HeJ, and BALB/cByJ. The WGS
            reads were placed uniquely on the MGSCv3 C57BL/6J assembly and SNP
            detection was carried out by SSAHA-SNP. 225,000 reads were
            annotated
            as STSs and 81,000 SNPs were annotated with alleles from C57BL/6J
            and the strain from which the particular read came. The validation
            rate for these SNPs was estimated at approximately 98%.
FEATURES   Location/Qualifiers
     source   1..802
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /strain="CZECHII/Ei"
                     /db_xref="taxon:10090"
                     /map="-11 22-762 26487420-26486680"
                     /clone_lib="CZECHII/Ei"
     STS   <1..>802
BASE COUNT   208 a   156 c   217 g   221 t
ORIGIN

Query Match   79.0%; Score 15.8; DB 11; Length 802;
  Best Local Similarity   89.5%; Pred. No. 7.1e+02;
  Matches   17; Conservative   0; Mismatches   2; Indels   0; Gaps   0;

Qy   2 ACCGACGCCTAATAGTTAAG 20
              || ||||| ||||||||||||
Db   52 ACAGACGCCAATAGTTAAG 70

RESULT 12
AC002108
LOCUS   AC002108   41125 bp   DNA   linear   ROD 29-MAY-1997
DEFINITION Genomic sequence from Mouse 4, complete sequence.
ACCESSION   AC002108
VERSION   AC002108.1 GI:2133882
KEYWORDS   HTG.
SOURCE   Mus musculus (house mouse)
  ORGANISM Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 41125)
  AUTHORS   Hawkins,T.L., Reeve,M.P., Christoffersen,A., Birren,B.W.,
            Fasman,K.H. and Lander,E.S.
  TITLE   Genomic sequence from Mouse 4
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 41125)
  AUTHORS   Hawkins,T.L., Reeve,M.P., Christoffersen,A., Birren,B.W.,
            Fasman,K.H., Lander,E.S., McKernan,K., Munro,C., Richardson,P.,
            Barna,N., Chang,A., Cooke,P., Daly,M.J., Forrest,C., Fripp,W.J.,

```
          Gage,D., Geraigery,K., Hagos,B., Jacotot,L., Lane,M., MacKenzie,J.,
          Marquis,N., McDermott,J., Moloney,N., Morrow,J., Nachman,A.,
          Naylor,J., O'Connor,T., Peterson,K., Rollins,G., Spencer,J.,
          Stilwell,J., Stone,C., Strickland,C., Sydney,K., Wilmer,F. and
          Zody,M.
  TITLE   Direct Submission
  JOURNAL Submitted (13-MAY-1997) Whitehead Institute/MIT Center for Genome
          Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE 3  (bases 1 to 41125)
  AUTHORS Hawkins,T.L., Reeve,M.P., Christoffersen,A., Birren,B.W.,
          Fasman,K.H., Lander,E.S., McKernan,K., Munro,C., Richardson,P.,
          Barna,N., Chang,A., Cooke,P., Daly,M.J., Forrest,C., Fripp,W.J.,
          Gage,D., Geraigery,K., Hagos,B., Jacotot,L., Lane,M., MacKenzie,J.,
          Marquis,N., McDermott,J., Moloney,N., Morrow,J., Nachman,A.,
          Naylor,J., O'Connor,T., Peterson,K., Rollins,G., Spencer,J.,
          Stilwell,J., Stone,C., Strickland,C., Sydney,K., Wilmer,F. and
          Zody,M.
  TITLE   Direct Submission
  JOURNAL Submitted (29-MAY-1997) Whitehead Institute/MIT Center for Genome
          Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT   On May 29, 1997 this sequence version replaced gi:2085725.
          The Staden databases, finishing information, and all
          chromatographic files used in the assembly of this clone are
          available from our anonymous ftp site.

          All repeats were identified using RepeatMasker: Smit, A.F.A. &
          Green, P. (1996-1997)
          http://ftp.genome.washington.edu/RM/RepeatMasker.html.
FEATURES          Location/Qualifiers
     source       1..41125
                  /organism="Mus musculus"
                  /mol_type="genomic DNA"
                  /db_xref="taxon:10090"
                  /chromosome="4"
                  /clone="Pla2s"
                  /clone_lib="unknown"
     repeat_region complement(6..528)
                  /rpt_family="ORR1A-INT"
     repeat_region complement(533..822)
                  /rpt_family="ORR1A2"
     repeat_region complement(1491..1581)
                  /rpt_family="PB1D7"
     repeat_region complement(2637..2990)
                  /rpt_family="MTE"
     repeat_region complement(3021..3154)
                  /rpt_family="B1-F"
     repeat_region complement(3301..3507)
                  /rpt_family="MER65_internal"
     repeat_region 3665..3706
                  /rpt_family="B4A"
     repeat_region complement(3736..4101)
                  /rpt_family="MTA"
     repeat_region complement(4102..4873)
                  /rpt_family="MT-INTERNAL"
     repeat_region complement(4878..5635)
                  /rpt_family="L1_MM"
     repeat_region complement(5636..6026)
                  /rpt_family="MT-INTERNAL"
     repeat_region complement(6027..6427)
                  /rpt_family="MTA"
     repeat_region 6535..6594
                  /rpt_family="(CA)n"
     repeat_region complement(6597..6760)
                  /rpt_family="MER65_internal"
     repeat_region complement(6797..6838)
                  /rpt_family="(CAT)n"
     repeat_region complement(8038..8366)
                  /rpt_family="RSINE2"
     repeat_region complement(9234..9305)
                  /rpt_family="ID5"
     repeat_region 11642..11709
                  /rpt_family="(CA)n"
     repeat_region complement(12447..12518)

     repeat_region /rpt_family="(GAA)n"
                  complement(12530..12874)
     repeat_region /rpt_family="RSINE2"
                  13468..13513
     repeat_region /rpt_family="(GAAAA)n"
                  14999..15141
     repeat_region /rpt_family="B1_MM"
                  15184..15384
     repeat_region /rpt_family="MLT1C"
                  16172..16294
     repeat_region /rpt_family="U1"
                  17852..17893
     repeat_region /rpt_family="(CA)n"
                  17894..17952
     repeat_region /rpt_family="(GA)n"
                  complement(18097..18167)
     repeat_region /rpt_family="ID2"
                  complement(18205..18266)
     repeat_region /rpt_family="(GAAA)n"
                  complement(18273..18782)
     repeat_region /rpt_family="Lx3"
                  complement(18812..18906)
     repeat_region /rpt_family="B2"
                  complement(18924..18963)
     repeat_region /rpt_family="B2"
                  complement(18964..19020)
     repeat_region /rpt_family="AT_rich"
                  complement(19034..19205)
     repeat_region /rpt_family="B2"
                  complement(19262..19390)
     repeat_region /rpt_family="B2"
                  complement(19626..19712)
     repeat_region /rpt_family="L1"
                  complement(19796..19958)
     repeat_region /rpt_family="MTD"
                  20077..20222
     repeat_region /rpt_family="RSINE1"
                  20236..20261
     repeat_region /rpt_family="(TAAA)n"
                  20448..20541
     repeat_region /rpt_family="MTE"
                  20622..21027
     repeat_region /rpt_family="MTD"
                  21063..21180
     repeat_region /rpt_family="ORR1C"
                  21933..21986
     repeat_region /rpt_family="(TAAAA)n"
                  complement(22006..22052)
     repeat_region /rpt_family="RMER1"
                  complement(22044..22236)
     repeat_region /rpt_family="B2"
                  complement(22237..22576)
     repeat_region /rpt_family="RMER1"
                  23096..23312
     repeat_region /rpt_family="B3"
                  23418..23493
     repeat_region /rpt_family="(GA)n"
                  complement(23624..23718)
     repeat_region /rpt_family="(GGGA)n"
                  complement(23890..23924)
     repeat_region /rpt_family="GC_rich"
                  complement(24935..25222)
     repeat_region /rpt_family="RSINE2"
                  complement(26023..26189)
     repeat_region /rpt_family="RSINE1"
                  26329..26373
     repeat_region /rpt_family="(TAGA)n"
                  complement(26375..26532)
     repeat_region /rpt_family="LIMA4A"
                  complement(27477..27535)
     repeat_region /rpt_family="LINE2"
                  27800..28003
     repeat_region /rpt_family="B3"
```

```
repeat_region        /note="3' end of clone 296C9"
                     1..941
                     /note="5PART"
                     /rpt_family="L1PA15"
repeat_region        955..1689
                     /rpt_family="L1ME1"
repeat_region        1690..1818
                     /rpt_family="L1ME1"
repeat_region        2730..2819
                     /rpt_family="(TA)n"
repeat_region        complement(2760..2844)
                     /rpt_family="(TAGA)n"
repeat_region        complement(2919..2949)
                     /rpt_family="(CATA)n"
repeat_region        complement(3023..3082)
                     /rpt_family="(TAGA)n"
repeat_region        3571..9708
                     /rpt_family="L1PA2"
repeat_region        complement(12194..12265)
                     /rpt_family="MIR"
repeat_region        13167..13663
                     /rpt_family="L1MC/D"
repeat_region        13664..13860
                     /rpt_family="THE1B"
repeat_region        13885..13933
                     /rpt_family="THE1A"
repeat_region        13934..15507
                     /rpt_family="THE1A-internal"
repeat_region        15508..15860
                     /rpt_family="THE1B"
repeat_region        complement(16262..16482)
                     /rpt_family="MIR"
repeat_region        complement(17454..17663)
                     /rpt_family="MIR"
repeat_region        17767..17996
                     /rpt_family="MIR"
repeat_region        18154..18180
                     /rpt_family="LTR34"
repeat_region        18181..18475
                     /rpt_family="LTR34"
repeat_region        18476..19159
                     /rpt_family="LTR34"
repeat_region        complement(19222..19853)
                     /rpt_family="MER4D"
repeat_region        complement(19856..19918)
                     /rpt_family="MER4B"
repeat_region        complement(19919..20233)
                     /rpt_family="AluSx"
repeat_region        complement(20234..20428)
                     /rpt_family="MER4B"
repeat_region        complement(20733..20912)
                     /rpt_family="MER5A"
repeat_region        complement(21263..21299)
                     /rpt_family="MER49"
repeat_region        complement(21345..21539)
                     /rpt_family="MIR"
repeat_region        22945..23172
                     /rpt_family="LTR1"
repeat_region        23083..23214
                     /rpt_family="LTR28"
repeat_region        23247..23556
                     /rpt_family="LTR1"
repeat_region        24587..24661
                     /rpt_family="LOR1a"
repeat_region        complement(24784..24901)
                     /rpt_family="LTR28"
repeat_region        complement(24389..25536)
                     /rpt_family="LTR1"
repeat_region        2970..29969
                     /rpt_family="AluSx"
repeat_region        29970..29995
                     /rpt_family="(CAAA)n"
                     30166..30384
```

```
repeat_region        28019..28120
                     /rpt_family="PB1D7"
repeat_region        28170..28195
                     /rpt_family="AT_rich"
repeat_region        complement(28215..28356)
                     /rpt_family="B1_MM"
repeat_region        complement(28357..28377)
                     /rpt_family="B4A"
repeat_region        complement(29678..29778)
                     /rpt_family="B1-F"
repeat_region        complement(29806..29855)
                     /rpt_family="ID3"
repeat_region        complement(30229..30537)
                     /rpt_family="RSINE2"
repeat_region        complement(31817..31860)
                     /rpt_family="(CA)n"
repeat_region        complement(32109..32147)
                     /rpt_family="PB1D9"
repeat_region        complement(32139..32256)
                     /rpt_family="B4A"
repeat_region        32796..32990
                     /rpt_family="B4A"
repeat_region        33019..33091
                     /rpt_family="(CAGA)n"
repeat_region        complement(33438..33489)
                     /rpt_family="MIR"
repeat_region        34814..34946
                     /rpt_family="MER74"
repeat_region        34956..35035
                     /rpt_family="B4A"
repeat_region        35036..35158
                     /rpt_family="PB1"
repeat_region        35163..35261
                     /rpt_family="(GGGA)n"
repeat_region        complement(35364..35466)
                     /rpt_family="(GGGA)n"
```

```
Query Match          79.0%;  Score 15.8;  DB 10;  Length 41125;
Best Local Similarity 89.5%;  Pred. No. 6.7e+02;  Indels 2;  Gaps 0;
Matches 17;  Conservative 0;  Mismatches 0;

Qy       2 ACCGACGCTAATAGTTAAG 20
           |||||||||||||||||
Db   15279 ACAGACGCTAATAGTTATG 15297
```

```
RESULT 13
AF117830/c
LOCUS       AF117830        71516 bp    DNA     linear   PRI 13-JAN-1999
DEFINITION  Homo sapiens clone from 8q21.3, complete sequence.
ACCESSION   AF117830
VERSION     AF117830.1  GI:4151949
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 71516)
AUTHORS     Platzer,M. and Varon,R.
TITLE       Direct Submission
JOURNAL     Submitted (30-DEC-1998) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany
COMMENT     This sequence is part of a larger genomic contig. The start of this
            sequence is directed towards the centromere. The start of this
            sequence (1..250) overlaps with the end of the neighbouring
            Acc_number AF049895 (330001..330250).
FEATURES             Location/Qualifiers
     source          1..71516
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="8"
                     /map="8q21.3"
```

```
repeat_region    complement(68073..68278)
                 /rpt_family="MIR"
repeat_region    complement(68668..68963)
                 /rpt_family="AluSq"
repeat_region    69033..69229
                 /rpt_family="L2"
repeat_region    69488..69795
                 /rpt_family="AluSx"
repeat_region    complement(70513..70782)
                 /rpt_family="L1ME"

BASE COUNT    22610 a   13954 c   13537 g   21415 t
ORIGIN

Query Match              79.0%;   Score 15.8;  DB 9;   Length 71516;
Best Local Similarity   89.5%;   Pred. No. 6.7e+02;
Matches   17;   Conservative    0;   Mismatches    2;   Indels    0;   Gaps    0;

Qy      1 CACCGACGCTAATAGTTAA 19
          ||| |||||||| ||||||
Db  41897 CACAGACGCTAAGAGTTAA 41879


RESULT 14
ATT12H17
LOCUS          ATT12H17             89350 bp    DNA            linear   PLN 03-FEB-1998
DEFINITION     Arabidopsis thaliana DNA chromosome 4, BAC clone T12H17 (ESSAII
               project).
ACCESSION      AL021635
VERSION        AL021635.1   GI:2827538
KEYWORDS       .
SOURCE         Arabidopsis thaliana (thale cress)
  ORGANISM     Arabidopsis thaliana
               Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
               Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
               rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE      1
  AUTHORS      Bevan,M., Hilbert,H., Braun,M., Holzer,E., Brandt,A.,
               Duesterhoeft,A., Bancroft,I., Mewes,H.W., Mayer,K. and Schueller,C.
  JOURNAL      Unpublished
REFERENCE      2  (bases 1 to 89350)
  AUTHORS      EU Arabidopsis sequencing project.
  TITLE        Direct Submission
  JOURNAL      Submitted (03-FEB-1998) MIPS, at the Max-Planck-Institut fuer
               Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, Project
               Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge
               Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,
               E-mail: michael.bevan@bbsrc.ac.uk
FEATURES             Location/Qualifiers
     source          1..89350
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /variety="Columbia"
                     /db_xref="taxon:3702"
                     /chromosome="4"
     gene            669..1151
                     /gene="T12H17.10"
     CDS             669..1151
                     /gene="T12H17.10"
                     /note="similarity to auxin-induced protein 10A, Glycine
                     max., PIR2:JQ1099"
                     /codon_start=1
                     /product="predicted protein"
                     /protein_id="CAA16547.1"
                     /db_xref="GI:2827539"
                     /db_xref="SPTREMBL:O49643"
                     /translation="MKRLRGKIGRHPVKLFKWITRSRRNQTGRKSRQCLTGLILNPVTKI
                     YSLARRCLRRGANRLCGGMKPGQTRLGNEPXTPTVPRGHLVVHVGESGEDTRRVVVPV
                     IYFNHPLFGSLLEQAERVVGFEQPGRIMIPCRVSDFEKVQMRIAAWDHCRRKSTFXIL
                     "
     exon            669..1151
                     /gene="T12H17.10"
                     /number=1
     gene            2406..2983
```

```
repeat_region    /rpt_family="L2"
                 complement(30393..30585)
                 /rpt_family="MIR"
repeat_region    30797..32782
                 /rpt_family="L2"
repeat_region    32937..33070
                 /rpt_family="L2"
repeat_region    complement(35789..35953)
                 /rpt_family="L2"
repeat_region    complement(36340..36703)
                 /rpt_family="THE1C"
repeat_region    37049..37878
                 /rpt_family="L2"
repeat_region    38812..38834
                 /rpt_family="POLY_A"
repeat_region    39226..39374
                 /rpt_family="MIR"
repeat_region    39375..39559
                 /rpt_family="MER45B"
repeat_region    39600..39698
                 /rpt_family="MER45C"
repeat_region    39758..40022
                 /rpt_family="MER45C"
repeat_region    40751..41255
                 /rpt_family="L2"
repeat_region    complement(42153..42241)
                 /rpt_family="L2"
repeat_region    42260..42399
                 /rpt_family="MER3"
repeat_region    42747..42799
                 /rpt_family="MIR"
repeat_region    44670..45617
                 /rpt_family="L1MB8"
repeat_region    47452..47479
                 /rpt_family="POLY_A"
repeat_region    48024..48134
                 /rpt_family="MER5A"
repeat_region    48218..48380
                 /rpt_family="L2"
repeat_region    50555..51031
                 /rpt_family="LTR22"
repeat_region    complement(51301..51541)
                 /rpt_family="MIR"
repeat_region    51765..51311
                 /rpt_family="THE1C"
repeat_region    52272..52772
                 /rpt_family="MLT1D"
repeat_region    complement(55159..55184)
                 /rpt_family="(CA)n"
repeat_region    complement(59095..59237)
                 /rpt_family="(TGGA)n"
repeat_region    59261..59434
                 /rpt_family="MIR"
repeat_region    62113..62575
                 /rpt_family="L2"
repeat_region    complement(65342..65488)
                 /rpt_family="MIR"
repeat_region    65602..65900
                 /rpt_family="L2"
repeat_region    65925..66258
                 /rpt_family="MLT1A1"
repeat_region    66304..66346
                 /rpt_family="(TAAA)n"
repeat_region    67132..67177
                 /rpt_family="(CAAAC)n"
repeat_region    67178..67305
                 /rpt_family="FLAM_A"
repeat_region    67306..67319
                 /rpt_family="(CAAAC)n"
repeat_region    67388..67713
                 /rpt_family="MLT1B"
repeat_region    67738..68071
                 /rpt_family="AluJo"
```

```
CDS             /gene="T12H17.20"
                join(2406..2754,2844..2983)
                /note="similarity to predicted proteins, Arabidopsis
                thaliana"
                /codon_start=1
                /product="predicted protein"
                /protein_id="CAA16548.1"
                /db_xref="GI:2827540"
                /db_xref="SPTREMBL:O49644"
                /translation="MAYTNKVAVAVGAAVVFLAVVMNPRWTEAQTYPKLDRLCVMMIP
                DILEECFTHDRLKPTEDCCNDLKNATMTQVDCLCDNFLESLSFSDLSRIFSAGVLKKC
                DVSHKYMCQAAKNRGEAKGGRNSTTCDNSITNTSVGGKNKVATSMSAFGLVAILLFV
                MF"

exon            2406..2754
                /gene="T12H17.20"
                /number=1

intron          2755..2843
                /gene="T12H17.20"
                /number=1

exon            2844..2983
                /gene="T12H17.20"
                /number=2

tRNA            3327..3400
                /gene="T12H17.30"
                /product="tRNA-Asn"
                /note="codon recognized: AAC"

gene            4696..5489
                /gene="T12H17.30"

CDS             join(4696..5035,5407..5489)
                /gene="T12H17.30"
                /note="contains EST gb:N38564"
                /codon_start=1
                /product="putative protein"
                /protein_id="CAA16549.1"
                /db_xref="GI:2827541"
                /db_xref="SPTREMBL:O49645"
                /translation="MAYTNKVTISAAVATMMLFLAVTIVDAQSMPPMKFNPVCALAD
                LPNIVQLCYFNLDLTPSEECCNDLKSSSTIQVNCLCDNFIAHPSNGNISQARYDLVNS
                ACGVADKPACKGGDASGSGSTNKIAASMVLLGLVASLFF"

exon            4696..5035
                /gene="T12H17.30"
                /number=1

intron          5036..5406
                /gene="T12H17.30"
                /number=1

exon            5407..5489
                /gene="T12H17.30"
                /number=2

gene            6595..7105
                /gene="T12H17.40"
                /number=1

CDS             join(6595..6940,7035..7105)
                /gene="T12H17.40"
                /codon_start=1
                /product="hypothetical protein"
                /protein_id="CAA16550.1"
                /db_xref="GI:2827542"
                /db_xref="SPTREMBL:O49646"
                /translation="MAVAVTAVLFLAVVIAPQWTETKKPPRFSDTSDTSGTSGRDRRT
                MCPLSIPGIVQNCYATLNAFPSKECCKDLKTASKREVTCLCNNVIAHPDPLYNTNQV
                NKACGVLDKYACDVAPTFIVYTNYDMIIIAVIQDP"

exon            6595..6940
                /gene="T12H17.40"
                /number=1

intron          6941..7034
                /gene="T12H17.40"
                /number=1

exon            7035..7105
                /gene="T12H17.40"
                /number=2

gene            8019..9209
                /gene="T12H17.50"

gene            complement(8019..9209)
                /gene="T12H17.50"

CDS             complement(8019..9209)
                /gene="T12H17.50"
                /note="similarity to predicted protein, Arabidopsis
                thaliana"
                /codon_start=1
                /product="predicted protein"
                /protein_id="CAA16551.1"
                /db_xref="GI:2827543"
                /db_xref="SPTREMBL:O49647"
                /translation="MEKNQNPNTWSDLPLDLLNLJVFKRLSFANFRQAKSVCSSWYSAS
                KQSVPKNQIPWLMLFPKDKNNNKNSSCTIFFNPERKDQLYQTQDLGVEFAKSVCLATY
                GSWLIMQDSKYNLYILNPFTYEKIGLPAIRSQQVGMVKVDQTIDDFLFDDHNHVKL
                FKGNNTVRTPVFWIDEKTKEYIALWGLGYWCVVYAKNGDKLWNQIPEIILDSLDMVYK
                DHKLYSFSYRNLFTILDFSGEIPRKAFQCFMHVRSEWLSPRSRQLSNSWCVAETKLV
                VTVTGDVLLVERMLRHWSRIQSFNVYKFYSSGTFLDKYELADSLGDEAMLLDLGITVL
                ANEVEGLHRNTIYFSDSHDTTTKDLFLFNVETREMEPLHKFDFDCSLLELSARWFLPS
                FSHT"

exon            complement(8019..9209)
                /gene="T12H17.50"
                /number=1

gene            9990..13709
                /gene="T12H17.60"

CDS             join(9990..10023,10239..10535,10564..10664,10910..11292,
                11755..11809,11898..12053,12143..12311,12391..12455,
                12534..12674,12769..12885,13024..13096,13244..13380,
                13468..13525,13612..13709)
                /gene="T12H17.60"
                /note="similarity to p48, Homo sapiens, PATCHX:G904032
                contains EST gb:Aa605571, H76919, H76638, R30271, Z25650"
                /codon_start=1
                /product="HSP associated protein like"
                /protein_id="CAA16552.1"
                /db_xref="GI:2827544"
                /db_xref="SPTREMBL:O49648"
                /translation="MRRTWVLAEESSVAIAPLLAEPQSTMFPEMTPECATVMPDLLEK
                CFATGSVTPTEDCCTDLKSATSTQVTCLCDNYIANPAVSNITGPYSKAITTKCGVFDK
                YSCDGTSKVISNKNGAFMVINGFWFCVVKEKKRKEEAAVAMGVLTLLGLINGLIVAH
                GKTRQQNKPVCKLLFQFQHFGFCSIRFSANKKARSRSKTGDAYPFRNEKVLEQKPYS
                LPFVPQILEIETRLKTKTKRLLTLRPKPTLFLFLSLSRRSEADFSQKDGFNEAYLGAK
                IPTGVHEEDKDTKPRSFVVEESDDDMDETEEVKPKVEEEEEDEIVESDVELEGDTVE
                PDNDPPQKMGDSSVEVTDENREAAQEAKGKAMEALSEGNFDEAIEHLTRAITLNPTSA
                IMYGNRASVYIKLKKPNAAIRDANAALEINPDSAKGYKSRGMARAMLGEWAEAAKDLH
                LASTIDYDEEISAVLKKVEPNAHKLEEHRRKYDRLRKEREDKKAERDRLRRRAEAQAA
                YDRAKKEEQSSSRPSGGGFPGGMGGGMPGMGGGMPAGMGGGMPGAGGGMPGGGGMP
                GGMDFSKILNDPELMTAFSDPEVWAALQDVMKNPANLAKHQANPKVAPVIARVMGKFA
                GPQ"

exon            9990..10023
                /gene="T12H17.60"
                /number=1

intron          10024..10238
                /gene="T12H17.60"
                /number=1

exon            10239..10535
                /gene="T12H17.60"
                /number=1

intron          10536..10563
                /gene="T12H17.60"
                /number=2

exon            10564..10664
                /gene="T12H17.60"
                /number=2

intron          10665..10909
                /gene="T12H17.60"
                /number=3

exon            10910..11292
                /gene="T12H17.60"
                /number=4

intron          11293..11754
                /gene="T12H17.60"
                /number=4

exon            11755..11809
                /gene="T12H17.60"
                /number=5
```

intron   11810..11897
                /gene="T12H17.60"
                /number=5
exon   11898..12053
                /gene="T12H17.60"

Query Match   79.0%; Score 15.8; DB 8; Length 89350;
Best Local Similarity   89.5%; Pred. No. 6.7e+02;
Matches   17; Conservative   0; Mismatches   2; Indels   0; Gaps   0;

Qy   1 CACCGACGCTAATAGTTAA 19
          |||||||| ||||||||||
Db 18364 CACCGACTCTAATAGTTGA 18382

RESULT 15
AL844490/c
LOCUS   AL844490   108612 bp   DNA   linear   ROD 19-SEP-2002
DEFINITION Mouse DNA sequence from clone RP23-98P13 on chromosome X, complete
            sequence.
ACCESSION   AL844490
VERSION   AL844490.4 GI:23304656
KEYWORDS   HTG.
SOURCE   Mus musculus (house mouse)
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 108612)
AUTHORS   Ellwood,M.
TITLE   Direct Submission
JOURNAL   Submitted (19-SEP-2002) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk
            On Sep 23, 2002 this sequence version replaced gi:22798535.
            --------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk

COMMENT   During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-98P13 is
            from the RPCI-23 Mouse PAC Library
            constructed by the group of Pieter de Jong.
            For further details see http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6.
FEATURES   Location/Qualifiers
     source   1..108612
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="X"
                     /clone="RP23-98P13"
                     /clone_lib="RPCI-23"
BASE COUNT   33150 a   21828 c   21230 g   32404 t
ORIGIN

Query Match   79.0%; Score 15.8; DB 10; Length 108612;
Best Local Similarity   89.5%; Pred. No. 6.6e+02;
Matches   17; Conservative   0; Mismatches   2; Indels   0; Gaps   0;

Qy   1 CACCGACGCTAATAGTTAA 19
          |||||||| ||||||||||
Db 31293 CACAGACGCTATTAGTTAA 31275

Search completed: November 17, 2003, 22:06:03
Job time : 597.3 secs